AOJP DATE 28/11/96 PCT NUMBER PCT/GB96/00700

AU9651169

(PCT)

(51) International Patent Classification 6 : G01W 1/10, G09B 29/00

A1

(11) International Publication Number: WO 96/29619

(43) International Publication Date: 26 September 1996 (26.09.96)

(21) International Application Number: PCT/GB96/00700

(22) International Filing Date: 18 March 1996 (18.03.96)

(30) Priority Data:
9505387.2 17 March 1995 (17.03.95) GB

(71) Applicant (*for all designated States except US*): UNIVERSITY OF LEEDS [GB/GB]; Leeds, West Yorkshire LS2 8JT (GB).

(72) Inventors; and
(75) Inventors/Applicants (*for US only*): LENNON, Jack [GB/GB]; 12 Wellhouse Avenue, Leeds, West Yorkshire LS8 4BY (GB). TURNER, John, Richard, George [GB/GB]; 58 Ash Hill Drive, Leeds, West Yorkshire LS17 8JP (GB).

(74) Agent: STANLEY, David, William; The Innovation Centre, University Road, Heslington, York Y01 5DG (GB).

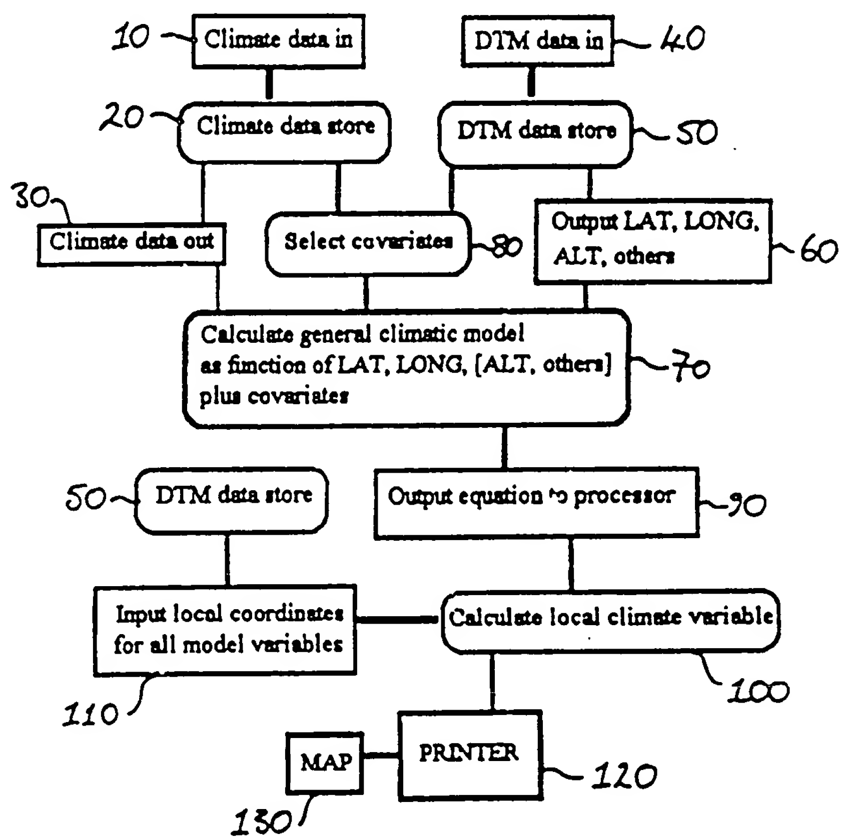
(81) Designated States: AL, AM, AT, AU, AZ, BB, BG, BR, BY, CA, CH, CN, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IS, JP, KE, KG, KP, KR, KZ, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO patent (KE, LS, MW, SD, SZ, UG), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: MAPS

(57) Abstract

Apparatus for drawing a contour map of a spatial area (e.g. a country) comprises means (10, 20) for recording a set of first data values (e.g. temperature), each measured at a respective predetermined point in the spatial area. Means (40, 50) is also provided for recording, for each of the predetermined points, various sets of further data values (e.g. latitude, longitude, altitude). Means (70) for fitting a mixed spline-regression model to the set of first data values is further provided, using as a spline variable at least one set of a plurality of sets of the further data values that have been selected by a multiple regression analysis as predictive of the set of first data values, and using the others of the selected sets of further values as covariates. Further means (100) predicts, from the model, values of the first data at a plurality of points in the spatial area; and there is provided means (120), using the predicted values, for drawing a map of the spatial area, in which the predicted values are depicted. Using such apparatus, contour maps may readily be drawn for all points, predictive to a high degree of accuracy of a desired variable (e.g. temperature) which has been measured only at some points.



MAPS

This invention relates to the production of maps and is concerned particularly, although not exclusively, with the production of maps showing spatial distributions of climate - for example, the spatial distribution of temperature over a geographical area.

The possibility of imminent climatic change has focused attention on the climate and its effects: it has become highly desirable to describe and predict climates. Many aspects of the distribution and functioning of organisms, with important implications for both ecology and agriculture, may be strongly influenced by climatic factors. In this context, it may be wished to estimate the climate from a place or area where there is no meteorological data. Preferred embodiments of the present invention are concerned with the prediction of mean monthly temperature from place to place, given that, inevitably, there is a limit to the number of climate recording stations at which temperature and other climatic variables may be recorded.

There has been doubt and controversy over the best method for producing a map of a complete climatic surface from a limited set of stations. Attempts have been made to use the standard statistical procedure of multiple regression to predict assorted climatic variables including temperature from a set of topographic and location variables, but these have been criticised and have found little favour in Britain. A more generally used method involves some sort of interpolation to fit a surface between

recorded points; traditionally this has consisted of drawing isopleths on a map by hand.

Recently, workers in Australia and New Zealand have used a thin plate spline method of Wahba G and Wendelberger J *(1980, Some new mathematical methods for variational objective analysis using splines and cross validation, Monthly Weather Review 108, 1122-1143)* to predict the spatial distribution of temperature and rainfall, and have extensively applied the calculated surfaces to explanations of individual species distributions - for example, Hutchinson MF *(1989, A new objective method for spatial interpolation of meteorological variables from irregular networks applied to the estimation of monthly mean solar radiation, temperature, precipitation and windrun. In Fitzpatrick EA and Kalma JD Need for climatic and hydrological data in agriculture in Southeast Asia. Proceedings of the United Nations University Workshop, December 1983. CSIRO Canberra, Division of Water Resources Technical Memorandum 89/5, 95-104).*

The National Environment Research Council recently opened the extensive TIGER IV programme (Terrestrial Initiative in Global Environmental Change) to investigate a wide range of questions as to how climate is related to ecology and conservation. The work of the present inventors on the climatic relations of the British fauna (Turner JRG, Gatehouse CM and Corey CA *(1987, Does solar energy control organic diversity? Butterflies, moths and the British climate, Oikos 48, 195-203)*; Turner JRG, Lennon JJ and Lawrenson JA *(1988, British bird species distributions and the energy theory, Nature 335, 539-541)*), and the proliferation of such projects under TIGER IV, has rendered urgent the construction of a set of maps of

the British climate. Maps on a suitable fine scale are generally not available. The officially published maps show monthly temperature by hand-interpolated isopleths *(Meteorological Office, 1975, Maps of mean and extreme temperature over the United Kingdom 1941-1970, HMSO, London)*: prediction at intermediate points has to be by guesswork. In addition, the temperature maps are corrected to sea level, which is meteorologically convenient because removing the overwhelming effect of altitude shows the underlying trends. Unfortunately, this renders the maps ecologically meaningless. Indeed a surprising number of ecologists have failed to realise that published temperature maps are drawn at sea level, and have consequently tried to match species distributions against these fictional isopleths, with questionable results. Overlays of temperature in wild-life atlases are likewise always fictional sea-level maps. Plots of raw station data as colour-coded discs on a map are excellent for objectivity but very hard to interpret.

True temperature maps are indeed difficult to draw because of the need to superimpose an altitude map on a sea-level isotherm map. In addition, because the rougher surface at real altitude can be interpolated much less reliably than the smoother temperature surface at sea level, it is not desirable to attempt interpolation between unadjusted station readings. Effectively, such a routine would be attempting to predict the topographic surface of the country from a thoroughly inadequate set of points. Only comparatively recently, fine grained digital terrain models (DTMs) have become available. Such DTMs give a detailed altitude map of a geographical area, may be produced by satellite surveying techniques, and are typically available as electronically stored digital data. The more detailed altitude data enables the effect of altitude on climatic maps to be better taken into

account. However, real climatic conditions are clearly affected by more than simple altitude alone, so a problem still remains in predicting such climatic conditions with accuracy.

According to one aspect of the present invention, there is provided apparatus for drawing a map of a spatial area, comprising:

means for recording a set of first data values, each measured at a respective one of a plurality of predetermined points in said area;

means for recording a plurality of sets of further data values, each data value pertaining to a respective one of said predetermined points;

means for fitting a mixed spline-regression model to the set of first data values, using as a spline variable at least one set of a plurality of sets of further data values that have been selected by a multiple regression analysis as predictive of said set of first data values, and using the others of said selected sets of further values as covariates;

means for predicting, from said model, values of said first data at a plurality of points in said spatial area; and

means, using said predicted values, for drawing a map of the spatial area, in which said predicted values are depicted.

According to another aspect of the present invention, there is provided a method of drawing a map of a spatial area, comprising the steps of:

(a) recording a set of first data values, each measured at a respective one of a plurality of predetermined points in said area;

(b) recording a plurality of sets of further data values, each data value pertaining to a respective one of said predetermined points;

(c) using those of said sets of further data values that have been selected by a multiple regression analysis as predictive of said set of first data values;

(d) fitting a mixed spline-regression model to the set of first data values, using at least one of said selected sets of further values as a spline variable, and using the others of said selected sets of further values as covariates;

(e) predicting, from said model, values of said first data at a plurality of points in said spatial area; and

(f) using said predicted values to draw a map of the spatial area, in which said predicted values are depicted.

A method as above may include a step (c1) of performing a multiple regression analysis on all of said sets of data values thereby to select those

of said plurality of sets of further data values as predictive of said set of first data values.

Preferably, in said step (c1) of performing a multiple regression analysis, at least one of said sets of further data values is rejected as non-predictive of said set of first data values.

Preferably, two of said sets of further data values comprise longitude and latitude values respectively and, in said step (d) of fitting a mixed spline-regression model to the set of first data values, only said latitude and longitude values are selected as spline variables.

Preferably, at least one of said sets of further data values is selected from the group comprising:

the altitude of each of said predetermined points above sea level;

the shortest distance from each of said predetermined points to the sea;

the maximum altitude to the east of each of said predetermined points in a ± 25 km north-south band; and

the variables listed in Table 1 below.

Preferably, the values of at least one of said sets of further values is measured or derived from a Digital Terrain Model (DTM).

Said first data values may be recorded over a first predetermined time period, and said map drawn to depict the values of said first data in a second time period.

Said first predetermined time period may comprise at least one full calendar year.

Said second time period may represent a predetermined time of year.

In the above, the term "map" may include a printed or displayed map, and/or a set or file of data from which a map may be constructed, and/or a look-up table of data from which various different maps may be derived.

For a better understanding of the invention, and to show how embodiments of the same may be carried into effect, reference will now be made, by way of example, to the accompanying diagrammatic drawings, in which:

Figure 1A is an outline map of Great Britain, showing the locations of "fitted" climate stations from which climatic data was used to derive a temperature map of the whole of Great Britain;

Figure 1B is an outline map of Great Britain, showing the locations of "check" climate stations which were used to check the accuracy of the temperature map derived from the fitted climate stations of Figure 1A;

Figure 2 is a graph to show "goodness of fit" (r^2) plotted against a percentage of fitted stations removed from a map drawing method;

Figure 3 is a temperature map, derived by one example of a method in accordance with the present invention, to show the average January temperature for the whole of Great Britain, to a resolution of 5 km;

Figure 4 is a temperature map similar to that of Figure 3, but showing variation between actual and predicted values;

Figure 5 is a map of the whole of Great Britain, derived by an example of a method in accordance with the present invention, to show growing seasons;

Figure 6 is a map of the whole of Great Britain, derived by an example of a method in accordance with the present invention, to show continentality; and

Figure 7 is a block schematic diagram of one example of apparatus embodying the invention, for drawing a map.

There now follows an example of the present invention, in which a temperature map of the whole of Great Britain is produced, to a 5 km resolution, from temperature data obtained from a relatively small number of scattered climate stations (weather stations), and from terrain data derived from a DTM of Great Britain. As will be shown, the accuracy of the temperature data of the map is better than 95%.

It is to be appreciated that the present invention is applicable to the production of maps other than temperature maps (e.g. rainfall maps, population maps), to the production of maps of areas other than Great Britain or other geographical areas (e.g. any spatial area of interest); and to maps at greater or lesser resolutions (e.g. with respect to Great Britain, accurate temperature maps down to a resolution of a few hundred metres may be possible). However, the following example of a temperature map of Great Britain is a useful one, and has been carried out experimentally.

In the following example, we have used a method which we have found to be an excellent predictor of temperature in Great Britain, and used it with a DTM to generate monthly, seasonal and other maps. We have used 30-year averages of monthly temperature from the national climate recording stations for the period 1941-1970, which is the last period for which data has been published *(Meteorological Office, 1976, Averages of temperature for the United Kingdom 1941-1970, HMSO, London)*, and we have prepared the maps on a 5 km grid. The more remote outlying islands (Orkney, Shetland, Isle of Man, Channel Isles, Scilly) were excluded because we were unable to obtain gridded topographic (DTM) data for use in regression analyses, as was Northern Ireland. This gave a total of 206 recording stations.

Briefly, we used the following procedure:

1. The stations were divided into two groups, designated as the "fitted" group (Figure 1A) and the "check" group (Figure 1B);

2. A model was built using the data from the fitted stations;

3. The goodness of fit of this model to the data from the check stations was calculated;

4. Data from all stations (that is the fitted and check stations combined) was then used to predict the climatic surface for the whole of Britain; and

5. Climatic maps were generated from this surface.

This procedure was performed to map a mean monthly temperature for each month of the year independently; seasonal and other data was then calculated by averaging, totalling or otherwise manipulating the monthly predictions.

The complete set of stations was divided into the two groups by first calculating a clustering index for each station (the sum of the reciprocal of the squared distances from all other stations $C_i = d_{ij}^{-2}$ for the ith station, where j sums the inter-station distance d across all remaining stations). The station with the maximum of this parameter was then removed and used as the foundation member of a new subset of stations; thereafter the clustering index of each station remaining within the first subset was calculated within each of the two subsets; the ratio of the two clustering indices was then calculated and the station with the largest value of this ratio was moved to the new subset. Application of this procedure to all stations in turn, produced two grids of stations with very similar overall geographical distributions; one of these was arbitrarily designated as the fitted group, and

the other as the check group. The distributions of the fitted and check stations are shown in Figures 1A and 1B.

The thin plate spline model of Wahba and Wendelberger (1980), developed for climate surfaces by Hutchinson (1989), fits a non-linear flexible *n*-dimensional surface to data points. The final surface is chosen by a routine which virtually removes each data point in turn, and estimates the deviation of its real value from the value predicted by the surface. The surface finally chosen is the one which achieves an optimum compromise between minimising the mean square deviation summed over all data points, and the overall roughness of the estimated surface. The surface can be fitted to latitude and longitude only, or may be fitted as a hyperdimensional surface to further variables such as altitude and distance from the sea. Spline surfaces may be calculated using M.F. Hutchinson's ANUSPLIN computer package.

In our work, we constructed a thin plate spline model from data which had been selected by a multiple regression method, and the resulting model is termed a mixed spline-regression model.

In order to construct regression models, we used a total of eighteen geographical and topographic variables, listed with abbreviations in Table 1 below, as a set of independent variables. The data was based on the national grid, sampled in 5 km squares from a DTM with points spaced at 500m. Within each 5 km square, the contained grid of 100 points was used to derive the mean, maximum and minimum altitude within the square (ALTME, ALTMA, ALTMI), the standard deviation of altitude within the

square (ALTSD), and the percentage of the square which is not sea surface (PLAND).

Table 1	
Abbreviation	Terrain variables derived from the digital terrain model
EAST	Longitude, as 4-figure Ordnance Survey grid reference (eastings)
NORTH	Latitude, as 4-figure Ordnance Survey grid reference (northings)
ALT	Altitude of meteorological station (meters)
ALTME (+1)*	Mean altitude of 100 points in 5 km square
ALTMA	Maximum altitude of 100 points in 5 km square
ALTMI (+1)	Minimum altitude of 100 points in 5 km square
ALTSD (+1)*	Standard deviation of altitude of 100 points in 5 km square
PLAND	Percentage of 5 km square that is land surface
ALTMEB (+1)*	Mean altitude of 8100 points in 45 km square
ALTMAB	Maximum altitude of 8100 points in 45 km square
ALTMIB (+1)*	Minimum altitude of 8100 points in 45 km square

ALTSDB (+1)*	Standard deviation of altitude of 8100 points in 45 km square
PLANDB	Percentage of 45 km square that is land surface
ALTWM	The maximum altitude to the west in a ± 25 km north-south band
ALTEM	The maximum altitude to the east in a ± 25 km north-south band
SLOS (+5)*	The large-scale slope of the land surface to the south (degrees)
SLOW (+5)*	The large-scale slope of the land surface to the west (degrees)
DIST	Shortest distance to the sea (km)
* for the log-linear regressions this value was added to each of the values to avoid problems with ln0.	

Additional variables were calculated for 45 km squares of the national grid (each consisting of 81 of the 5 km squares), by computing within each 45 km square the mean altitude, the average maximum altitude (mean of the 5 km maxima), the average minimum altitude (similarly), and the standard deviation within the 45 km square (root mean square deviation of the 8100 reference points within the square), and the percentage of the large square which is not sea-surface (names of variables as before, with the addition of 'B'). This set of variables allowed us to include coarse scale topographic information in the models.

The maximum land-height to both the east and west (ALTEM, ALTWM) of each 5 km square was found by searching the matrix of altitudes in a band 50 km wide (25 km north and 25 km south) extending east or west of the square as appropriate. These variables are intended to allow for possible effects of distant high ground, for example the formation of orographic cloud. The large scale slope of the ground in the vicinity of each 5 km square (SLOS, SLOW) was found as the difference in mean altitude between the reference square and the immediately adjacent 5 km square to the south or to the west. These slopes are small (at most ± one degree of arc) since the base line of 5 km is long in relation to the typical mean altitude.

The distance from the sea (DIST) was calculated from a digitised coastline. The proportion of the square (5 km or 45 km)(PLAND, PLANDB) that is land surface expresses, besides proximity to the sea, an abstraction of the shape of the coastline around that location. It was computed as the percentage of points at zero altitude lying outside the coastline (to exclude land surface at or below sea level).

In the derivation and verification of regression models, the actual grid east and grid north positions (to six figures) (EAST, NORTH) of the climate recording stations were used as variables. For other purposes, for example the construction of maps from mixed-spline regression models, the grid coordinates are those at the centre of the appropriate 5 km square. Similarly the altitude used in deriving and verifying regression and mixed spline-regression models was the true height of the station (ALT); in applications

of the regression and spline models this station height was replaced with the mean altitude of the 5 km square (ALTME).

The above variables were used in multiple regression equations of the form:

$$Tm = a + b_1x_1 + b_2x_2 + \ldots\ldots + b_nx_n,$$

where Tm = average monthly temperature, for a given month, and averaged over the sample period (thirty years);

a = a constant;

$b_1 \ldots\ldots b_n$ = regression coefficients

$x_1 \ldots\ldots x_n$ = variables

In the following, the variables x_1 .... x_n are the variables listed in Table 1, and explained above. The regression coefficients b_1 .... b_n vary for each month of the year, and will have a positive, negative or zero (or minimal) value. The following Table 2 gives the sign of the regression coefficients b_1 .... b_n for the twelve months of the year.

Table 2

MONTH	EAST	NORTH	ALT	DIST	PLAND	ALTME	ALTMA	ALTMI	ALTSD	ALTWM	ALTEM	PLANDB	ALTMEB	ALTMAB	ALTMIB	ALTSDB	SLOS	SLOW
JAN	-	-	-	+	+	.	+	-	-	.	.	-	+	.	-	-	-	-
FEB	-	-	-	+	-	-	+	-	.	.	.	-	.	+	-	-	-	[illegible]

MAR	-	-	-	+	-	-	+	-	.	.	+	.	.	+	-	-	-	.
APR	-	-	-	+	-	-	.	-	.	+	+	+	.	.	.	-	-	.
MAY	+	-	-	+	-	-	.	-	.	+	+	+	-	.	+	.	-	.
JUN	+	-	-	+	-	-	.	-	.	+	+	+	-	.	+	.	-	.
JUL	+	-	-	+	-	-	.	-	.	+	+	+	-	.	+	.	-	.
AUG	+	-	-	+	-	-	.	-	.	+	+	.	-	.	+	+	-	.
SEP	+	-	-	+	-	-	.	-	+	+	+	-	.	-	+	.	-	.
OCT	+	-	-	+	+	-	.	-	+	.	+	-	.	-	.	.	-	.
NOV	-	.	-	+	+		+	-	.	.	+	-	.	.	-	.	-	
DEC	-	-	-	-	+	-	+	-	.	.	.	-	.	+	-	-	-	

+ : variable included in model with positive regression coefficient.
- : variable included in model with negative regression coefficient.
. : variable not included in model (coefficient set at zero)

Within each set of variables the best predictive model in each month was chosen by using Statistical Analysis System routine PROC RSQUARE (SAS 1985). This system runs multiple regressions for all possible permutations of the independent variables and estimates the model with the largest r^2 for a given sized subset of independent variables. Thus for each month eighteen models were found, each model being the one with the largest r^2 out of the set of all models with the specified number of variables (between 1 and 18). For each month the model from this set of eighteen with the highest adjusted r^2 was selected; this in effect removed models which include variables that do not increase predictive power. Similar model selection procedures such as JP *(Judge GG, Griffiths WE, Hill RC, Lee T, 1980, The theory and practise of econometrics, Wiley, New York)*, PC *(Amemiya T, 1976, Selection of regressors, Technical Report # 225, Stanford*

University, Stanford, California) and Cp *(Mallows CL 1964 Some comments on Cp, Technometrics 15, 661-675)* were in strong agreement over which was the best model. This procedure was considered superior to stepwise elimination in which variables which are not individually significant can be dropped, since our procedure was concerned with the predictive ability of the model rather than with testing hypotheses about causal relationships between the dependent and independent variables *(Hocking RR 1976 The analysis and selection of variables in linear regression, Biometrics 32, 1-49)*.

In a mixed spline-regression model, the surface is fitted to some of the variables by the thin plate spline routine described above, but only after the surface has been corrected to its expected values derived from multiple regression on a further suite of variables (termed for this purpose *covariates*). Operationally, the parameters of regression on the covariates are computed, and the expected value of each data point is determined from the regression equation. These expected values are then treated as raw data to which the spline surface is fitted. In the present example we have treated a wide suite of variables derived from the DTM as covariates, leaving only latitude and longitude to be fitted by the thin plate spline. The covariates were selected from the DTM using multiple regression techniques.

There is a strong clustering of climate recording stations in the south of England, particularly in the London area. To examine the effect of such clustering on the regression models, stations in clusters were removed before the partition into the fitted and check datasets. Stations were removed from clusters in order of decreasing values of C_i (the sum of the reciprocal squared distances to the remaining stations - see above). The best predictive method

found (sLLr9 – Table 3) was then applied to the thinned out station grid; repetition of this procedure with an increasing percentage of stations removed allowed us to estimate the effect of clustering on the goodness of fit.

The accuracy of the fitted surfaces (i.e. fitted to the "fitted" stations) was estimated from the goodness of fit of the values which they predicted for the set of check stations compared with the actual meteorological readings at those stations. Fit was estimated as the coefficient of determination (r^2) of the correlation (r) between the observed and predicted values.

The r^2 for goodness of fit between the surface predicted from the fitted stations and the actual temperature values at the check stations are shown in Table 3 below. For ease of reference, regression methods are denoted by r, splines by s, with the variables following the letter: thus sLLrA denotes a spline surface fitted to latitude and longitude, with regression on altitude as a covariate. D denotes distance from the sea, and numerous variables are denoted by their number: r18 indicates regression on eighteen variables, with suffix f for the full set and s if a subset has been selected.

Table 3 Goodness of fit by cross-validation of tested methods of fitting monthly temperatures
Values are coefficients of determination

Abbrev	Method	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Average
rLLAD	regression four variables	850	871	889	882	843	840	899	924	921	888	849	829	874
rLLAlogD	regression four variables	864	874	884	885	855	857	907	924	919	888	861	845	[illegible]
rLLA	regression three variables	821	861	892	877	817	808	882	920	922	871	820	793	[illegible]

sLLrA	spline LAT, LONG regression ALT	914	904	898	890	901	895	929	932	922	920	910	907	91[illegible]
sLLrAD	spline LAT, LONG reg. ALT DIST	918	910	916	911	905	897	931	932	926	907	911	910	914
sLLr9	spline LAT,LONG reg 9 DTM covars	940	920	909	902	911	907	936	939	937	936	937	907	92[illegible]
rLLr9	regression LAT,LONG, 9 DTM vars	885	851	874	904	884	867	898	909	914	935	923	881	89[illegible]
r18s	reg. selection of 18 DTM vars	936	917	912	909	888	877	914	928	924	931	933	907	91[illegible]
r18f	regression all 18 DTM vars	866	870	893	907	872	862	894	904	905	923	923	864	8[illegible]
sLLr18f	spline LL reg. all 18 DTM vars	935	919	919	904	890	871	911	926	923	925	934	9[illegible]	91[illegible]
sLLr18s	spline LL reg selected DTM vars	940	921	919	907	91[illegible]	881	936	934	939	938	936	926	914

Abbreviations are explained in the text; also var. = variables, covars = covariates, reg. = regression. Values are r^2 for correspondence between actual and predicted values × 1000

In selecting covariates, multiple regression is markedly improved by the inclusion of distance from the sea, the improvement being slightly greater if this variable is transformed to its logarithm. The best overall regression model is that using a month by month selection from the eighteen DTM variables, selection being by the model building criteria described under Methods, with independent selection in each month: Table 2 above shows the significant monthly variables with the signs of their beta coefficients. This model shows an r^2 value exceeding 0.9 (equivalent to a correlation coefficient of at least 0.95) in most months of the year.

As would be expected, the pure multiple regression model noticeably diminishes in its predictive power if all eighteen DTM variables are included, with no attempt at selection (r18f). The regression model is also reduced in predictive power if a single set of the eleven consistent variables (the eleven variables that are included in the model in at least eight months of the year) is used in all months, instead of the model being built individually for each month (rLL9).

Regression models in which all independent variables, or all variables, were transformed to logarithms or square roots were almost identical with the equivalent untransformed models, and will not be discussed further.

The best overall fit is obtained from a mixed spline-regression model (sLLr9) using the nine DTM variables which had the greatest consistent performance in the regression model. That is, we have used the nine DTM variables (other than latitude and longitude) which are found from examination of Table 4 to have been selected for inclusion in the multiple regression model in at least nine months of the year. We have then fitted a thin plate spline with these nine variables as covariates and with latitude and longitude as the spline surface variables. This gives a set of models with every month having an r^2 of over 0.9. A somewhat worse fit is obtained if all sixteen DTM variables are included as covariates (sLLr16f). Building a separate mixed model for each month individually, using as covariates those DTM variables (excluding latitude and longitude) that were selected by the regression models for those individual months (so that for instance September used twelve covariates and October used only ten -- sLLr16f), produced no alteration in the average accuracy of prediction compared with the use of the nine consistent variables, the individually selected models achieving first place by a small margin in a majority of months, but giving a notably poor prediction for June.

According to a standard computation used in spline models, the accuracy of fit between the stations and the fitted spline surface has an r^2 better than 0.97 in all months of the year; such figures are frequently quoted to demonstrate the accuracy of thin plate splines. However, this figure

represents only the ability of the spline surface to conform to data points; it does not represent its ability to predict temperature in areas away from stations.

Figure 2 shows the effect on the fit of the best model (the mixed model sLLr9 just described), averaged over the whole year, produced by progressively removing stations from the grids of fitted and check stations. The stations have been removed in such a way that the most densely clustered stations are removed first, until only fifteen percent remain. Removal of up to seventy percent of temperature stations has little impact on prediction; the average annual r^2 shows only a small decline from its maximum value of 0.925. This suggests that about two-thirds of the most densely packed stations are redundant. Collapse of the method is then rapid. This suggests that satisfactory temperature surfaces could be produced by the mixed spline-regression method using only thirty percent of the fitted station grid, provided the selection was made carefully to exclude the most densely clustered stations. As the cross-verification procedure splits the station grid into two (the fitted and the check stations), this means that in fact only a half of thirty percent of the total 206 stations, that is a mere 32 well-spaced stations, are required for an adequate fit. Forty or fifty stations would perhaps give an adequate safety margin.

This thinning procedure has used the same set of nine consistent covariates throughout. It is possible that were the multiple regression model rerun independently at each stage of the thinning procedure, it would produce a slightly different recommended set of covariates. Thus some of the decreases in predictive power at intermediate percentages of stations

might be avoided; it is encouraging that at almost all stages of thinning as far as seventy percent removal, the average annual r^2 remains above 0.9.

The relationship between the clustering index of stations (C - see above) and their individual residual deviations from the predicted surface derived from the complete station set shows that, when the sign of the residual is taken into account, there is no relationship between clustering and the accuracy of prediction. The absolute value of the residual is significantly correlated with isolation in five months of the year: isolated stations are *more* accurately predicted. This effect probably occurs because a spline will conform more closely to isolated stations than to the individual stations in a dense cluster, where it can conform only to their average because of restrictions which are placed on the roughness of the spline surface. This might result in the method compromising its ability to predict temperatures at points remote from stations in areas where the stations are already sparse. We have shown that this is not the case - there is no significant relationship between the residual and the degree of isolation of the stations in any month; in other words, the method does not predict temperatures at points remote from the stations any better (or any worse) in areas with sparsely distributed stations. The spline thus conforms well to the best data available: the average value in a cluster of stations, and the individual values of isolated stations. There is no evidence that it sacrifices long distance accuracy at the cost of tight fits to closely spaced stations.

As mixed spline-regression models provided most satisfactory predictions of temperature, we derived a series of models that used all the British climate recording stations for the reference period: that is the fitted

and check station sets combined (certain outlying stations are still omitted as detailed above). The station clusters were not thinned out. A separate model was derived for each month, using the same nine consistent DTM variables in the mixed spline-regression method sLLr9.

Maps based on mixed spline-regression models were produced by using the model to fit a predicted value of the climatic variable to the suite of selected independent variables at each point on the 5 km grid. For most such points the variable ALT, which represents the altitude of the recording station, is meaningless, and it is therefore replaced for computation by ALTME, the mean altitude of the square, multiplied by the beta-coefficient of ALT.

A series of temperature maps was computed using the model which produced the overall best predictions: the mixed spline-regression method with the nine consistent DTM variables (sLLr9 -- Table 3). A map for January is plotted as Figure 3, to show the mean temperature for this month. The distributions for the four Manley quarters -- for instance a "spring" of March, April and May -- are almost identical with the respective cardinal months of January, April July and October. Figure 5 illustrates length of growing season, and Figure 6 continentality. We leave it to the reader's own judgement, what biological applications are suggested by these distributions, which in some cases appear to be the first to be published at real altitude (not corrected to sea level).

Seasonal and annual maps were calculated as the average of the appropriate monthly maps, with the months weighted according to their

lengths. Growing season was calculated as the number of days when the average temperature exceeded the index value of 6° Celsius, using linear interpolation of the annual temperature curve between the means of the relevant adjacent months. Degree days may be computed as the integral of the linearly interpolated annual temperature curve multiplying *days x daily temperature*. Continentality is computed as the difference between the mean temperatures of the warmest and coldest months, which are usually but not invariably July and January. All these statistics, that is growing season, degree days and continentality, were computed point by point on the 5 km grid.

As a check on possible biases in the final temperature surfaces, we have plotted month by month the geographical distributions of the residual differences between the real station values and the surface. We have found that there is a high level of consistency between the patterns exhibited throughout the year. There are no consistent overall geographical trends in the distribution of the residuals (Figure 4); this indicates that accuracy is not for instance biased by latitude, as it might have been if dense clusters had fitted markedly more or less easily than isolated stations, or towards the south of England, which might have been expected from the possible bias of the estimates of the beta coefficients of the regression model toward the values found in the region with the most densely clustered stations.

There is however some real information in the distribution of the largest residuals. Nearly all the large positive values (station warmer than predicted by model), and some of the weaker ones, are centred on urban areas. This effect is obviously explained by the well known formation of

heat islands in towns and cities. Those familiar with the map of Britain will be able to locate London, Southampton and Portsmouth, Brighton, Oxford, Nottingham, Birmingham, Liverpool and Manchester, Sheffield, Glasgow, Edinburgh, and Dundee in this way. The placing of the recording stations on the edge of some of the urban areas has weakened their effect, notably Manchester and Edinburgh, and the absence of a station in West Yorkshire (Leeds-Bradford) and Tyneside means that these conurbations cannot be detected. The heat island over London appears weak and fragmented apparently because the large number of stations within the urban area allows the predicted surface from the composite spline regression model (sLLr9) to rise consistently over this whole area (it is noteworthy that the comparable residuals from a pure DTM multiple regression model r18s which lacks this flexibility, show a major heat island on London). One quite strong and persistent positive residual occurs west of Birmingham, in a region of small rural towns; possibly this represents also the effect of interpolation between the individual small heat islands. There is no obvious comparable explanation for the distribution of negative residuals (station cooler than predicted by model), which generally occupy much larger and more diffuse areas than the positive residuals: possibly some of the most extreme negative values centre on stations that are in major frost hollows; this is for instance likely for Kielder Castle, which produces a moderate but consistent negative residual year-round in the Anglo-Scottish border, and which lies at the confluence of several steep narrow valleys. The model might therefore be improved by using even finer scale topographic data to investigate mesoclimatic effects in the vicinity of the recorder; clearly it could be improved by introducing an index of urbanisation as a covariate.

The largest residuals are less than one half of one degree, except for seven values which lie between 0.5 and 0.8 of a degree: this is consistent with the usual accuracy of thin plate splines when plotting temperature surfaces. Some of the effects on the accuracy of prediction of removing clustered stations (above) may indicate in addition or instead, the effect of removing urban stations, as there is a particularly large cluster in the London area.

The best fit of all was obtained by using a suite of DTM variables previously identified by multiple regression as having a significant contribution to the regression model. In this particular case, the year-round fit was about equally as good when the variables were selected by an independent model for each month, or when the nine variables that had the most consistent appearance over the whole year were used in all months. Although in this study we have used the slightly better year-round model for our final maps, we anticipate that in most applications the better result will be obtained from individual monthly models.

The predicted surface using the mixed regression-spline method is shown by cross-validation to have a better than ninety five percent correlation with reality in all months of the year.

Our maps of temperature (e.g. Figure 3) are among the first to be produced for Great Britain, in particular in attempting to plot at real altitude rather than reduced to sea level: our map of the length of the growing season appears to be the first such to cover the whole of Great Britain (previous attempts covered England and Wales only). Cross-validation

suggests that these maps are always better than 95 percent correlated with reality. It is likely that there is some geographical variation in accuracy, although it is difficult to say what pattern this might occur in, given that the maps of the residuals for temperature show no obvious trends apart from the effect of urbanisation. We suggest that clearly an index of urbanisation would improve temperature prediction. Examination of the DTM variables by graphical methods, and selective transformation to logarithmic and other scalings might produce some further improvement. The method we have employed here is clearly applicable to other meteorological data: minimum and maximum temperatures, either monthly or daily, would be easy to analyse in this way. A particularly desirable extension would be an analysis of rainfall patterns.

The maps have direct application to ecogeographical studies of data gathered predominantly in the period 1941-70, such as the *Atlas of the British Flora* (Perring and Walters 1962). However during the century up to the present in which climate data have been intensively studied it has been established that thirty-year averages provide quite accurate predictions for mean climate in subsequent periods. Our maps should therefore prove to be adequate for ecological studies that use data gathered on either side of this period.

The predictions in the above example are of a mesoclimate; detailed predictions of microclimate could require different considerations of local conditions, for instance in the prediction of frost-hollows or of the effects of local north and south slopes. Our predictions relate to the average altitudes of national grid squares of 5 km side; further local correction (using

for instance the standard adiabatic lapse rate) would be needed to predict the temperature of any point within such a square that deviated considerably from the average altitude.

Potential applications of climate surfaces of this type are clearly diverse, from studies of the limitation of individual species by climatic factors to studies of species richness. Our DTM mixed spline-regression model for temperature, gridded by interpolation to a resolution of 2 km, will be applied to the changing distribution of the British bird fauna and to explaining the distribution of bird diversity in the British winter. The general method used, fitting a thin plate spline surface with variables, derived from a DTM and selected by multiple regression techniques, as covariates, is likely to have wide applications in biological meteorology, general meteorology, ecology and geography; clearly it may be applicable to a wide variety of meteorological and other environmentally significant statistics.

For ecological purposes, mesoclimatic conditions of temperature at points distant from climate recording stations in Britain can be predicted with an accuracy of better than 95 per cent. The preferred method for plotting monthly temperature surfaces for Great Britain is a thin plate spline fitting the variables latitude and longitude, with the entry of selected DTM variables as covariates. Selection of the DTM variables is by multiple regression, using model-building criteria.

Although the above-described examples of the invention produce maps which may be, for example, either displayed on a visual display unit

(e.g. a monitor screen) or printed onto a print medium, other embodiments of the invention may alternatively produce data from which a map or table may be displayed or printed. For example, apparatus in accordance with certain embodiments of the invention may produce a file of electronic data that may be stored, e.g. in the form of a look-up table, the data representing, for example, temperature values that have been predicted by a method as described in the foregoing.

Figure 7 is a block schematic diagram of apparatus for drawing a map, which may operate in accordance with a method as described above.

The apparatus of Figure 7 comprises a main processor 100 for calculating a local climate variable, by predicting values of temperature at a plurality of points in the spatial area of interest (for example, the United Kingdom), using data input from a DTM data store 50, via an output device 110, and an equation generated by a processor 70 and output via an output device 90. The predicted values as calculated by the main processor 100 are fed to a printer 120, which outputs a printed map 130. For example, the printed map 130 may have the form shown in any of Figures 3 to 6.

The processor 70 calculates a general climatic model from climate data that is received from a climate data store 20 via an output device 30, and DTM data values that are output from a DTM data store 50 via an output device 60. A processor 80 receives both climate data from the store 20 and DTM data from the store 50, and selects therefrom covariates for the mixed spline-regression model that is calculated in the processor 70. The climate

data store 20 receives climate data from an input device 10, and the DTM data store 50 receives DTM data from an input device 40.

The apparatus of Figure 7 may be arranged to perform any of the methods described above, as examples of the present invention.

The reader's attention is directed to all papers and documents which are filed concurrently with or previous to this specification in connection with this application and which are open to public inspection with this specification, and the contents of all such papers and documents are incorporated herein by reference.

All of the features disclosed in this specification (including any accompanying claims, abstract and drawings), and/or all of the steps of any method or process so disclosed, may be combined in any combination, except combinations where at least some of such features and/or steps are mutually exclusive.

Each feature disclosed in this specification (including any accompanying claims, abstract and drawings), may be replaced by alternative features serving the same, equivalent or similar purpose, unless expressly stated otherwise. Thus, unless expressly stated otherwise, each feature disclosed is one example only of a generic series of equivalent or similar features.

The invention is not restricted to the details of the foregoing embodiment(s). The invention extends to any novel one, or any novel

combination, of the features disclosed in this specification (including any accompanying claims, abstract and drawings), or to any novel one, or any novel combination, of the steps of any method or process so disclosed.

CLAIMS:

1. Apparatus for drawing a map of a spatial area, comprising:

means for recording a set of first data values, each measured at a respective one of a plurality of predetermined points in said area;

means for recording a plurality of sets of further data values, each data value pertaining to a respective one of said predetermined points;

means for fitting a mixed spline-regression model to the set of first data values, using as a spline variable at least one set of a plurality of sets of further data values that have been selected by a multiple regression analysis as predictive of said set of first data values, and using the others of said selected sets of further values a· covariates;

means for predicting, from said model, values of said first data at a plurality of points in said spatial area; and

means, using said predicted values, for drawing a map of the spatial area, in which said predicted values are depicted.

2. A method of drawing a map of a spatial area, comprising the steps of:

(a) recording a set of first data values, each measured at a respective one of a plurality of predetermined points in said area;

(b) recording a plurality of sets of further data values, each data value pertaining to a respective one of said predetermined points;

(c) using those of said sets of further data values that have been selected by a multiple regression analysis as predictive of said set of first data values;

(d) fitting a mixed spline-regression model to the set of first data values, using at least one of said selected sets of further values as a spline variable, and using the others of said selected sets of further values as covariates;

(e) predicting, from said model, values of said first data at a plurality of points in said spatial area; and

(f) using said predicted values to draw a map of the spatial area, in which said predicted values are depicted.

3. A method according to claim 1 or 2, including a step (c1) of performing a multiple regression analysis on all of said sets of data values thereby to select those of said plurality of sets of further data values as predictive of said set of first data values.

4. A method according to claim 3, wherein, in said step (c1) of performing a multiple regression analysis, at least one of said sets of further data values is rejected as non-predictive of said set of first data values.

5. A method according to claim 2, 3 or 4, wherein two of said sets of further data values comprise longitude and latitude values respectively and, in said step (d) of fitting a mixed spline-regression model to the set of first data values, only said latitude and longitude values are selected as spline variables.

6. A method according to any of claims 2 to 5, wherein at least one of said sets of further data values is selected from the group comprising:

the altitude of each of said predetermined points above sea level;

the shortest distance from each of said predetermined points to the sea;

the maximum altitude to the east of each of said predetermined points in a ± 25 km north-south band; and

the variables listed in Table 1 above.

7. A method according to any of claims 2 to 6, wherein the values of at least one of said sets of further values is measured or derived from a Digital Terrain Model (DTM).

8. A method according to any of claims 2 to 7, wherein said first data values are recorded over a first predetermined time period, and said map is drawn to depict the values of said first data in a second time period.

9. A method according to claim 8, wherein said first predetermined time period comprises at least one full calendar year.

10. A method according to claim 8 or 9, wherein said second time period represents a predetermined time of year.

11. A method of drawing a map of a spatial area, substantially as hereinbefore described with reference to the accompanying drawings.

12. Apparatus for drawing a map of a spatial area, adapted to perform a method according to any of claims 2 to 11.

13. Apparatus for drawing a map of a spatial area, substantially as hereinbefore described with reference to the accompanying drawings.

14. Apparatus or a method for drawing a map of a spatial area, according to any of the preceding claims, wherein said first values represent temperature.

15. Apparatus or a method for drawing a map of a spatial area, according to any of the preceding claims, in combination with any or all of the features disclosed in this specification, including the accompanying abstract and drawings.

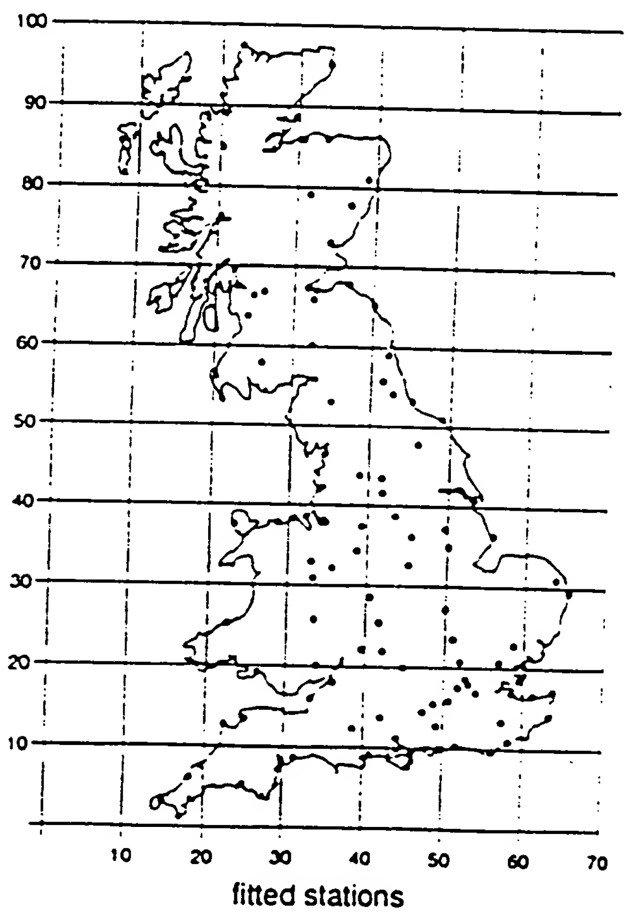

100
90
80
70
60
50
40
30
20
10
10 20 30 40 50 60 70
fitted stations

FIG.1A

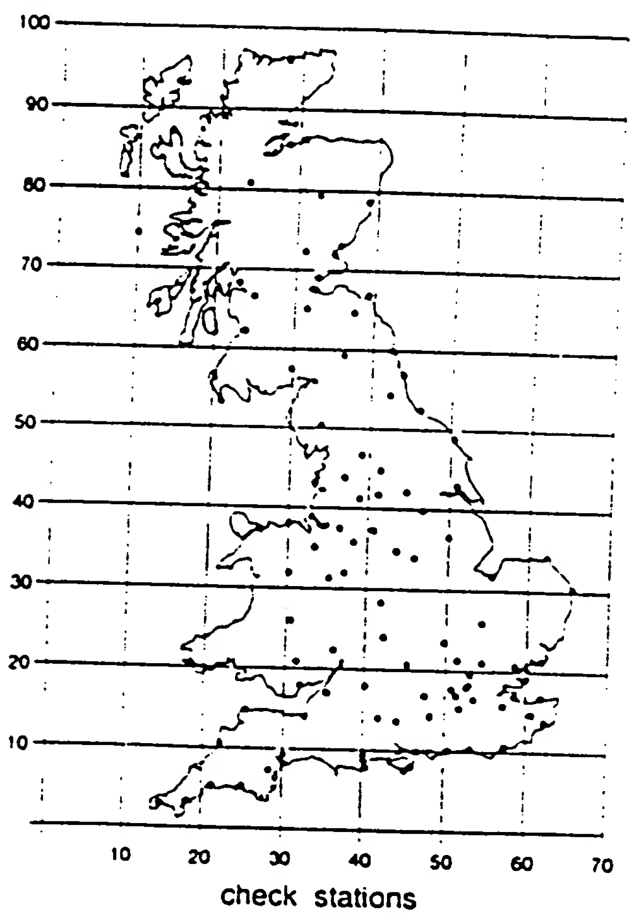

100
90
80
70
60
50
40
30
20
10
10 20 30 40 50 60 70
check stations

FIG.1B

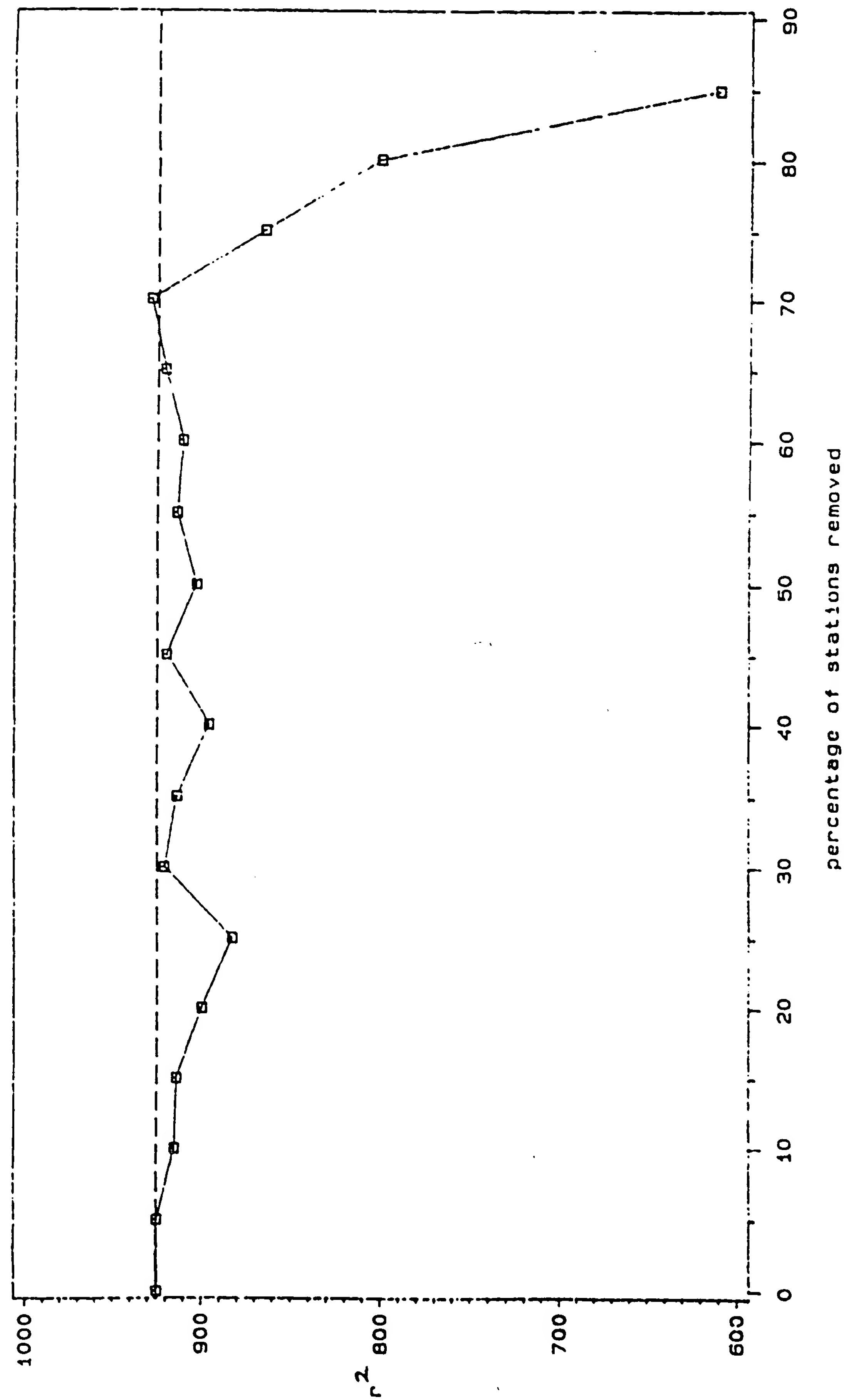
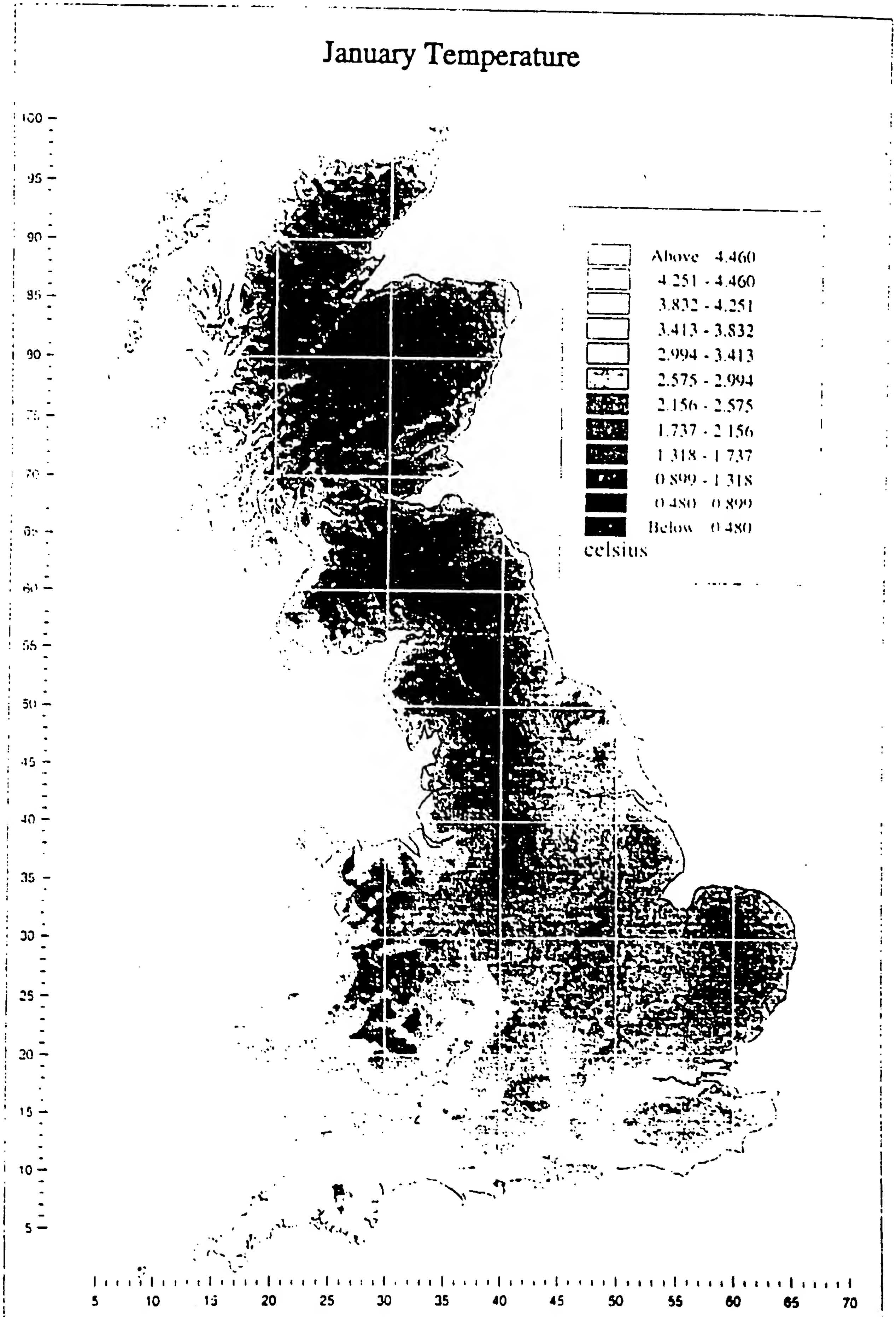

percentage of stations removed
0
10
20
30
40
50
60
70
80
90
r²
1000
900
800
700
600

Figure 3



January Temperature
Above 4.460
4.251 - 4.460
3.832 - 4.251
3.413 - 3.832
2.994 - 3.413
2.575 - 2.994
2.156 - 2.575
1.737 - 2.156
1.318 - 1.737
0.899 - 1.318
0.480 0.899
Below 0.480
celsius
5
10
15
20
25
30
35
40
45
50
55
60
65
70

Figure 4

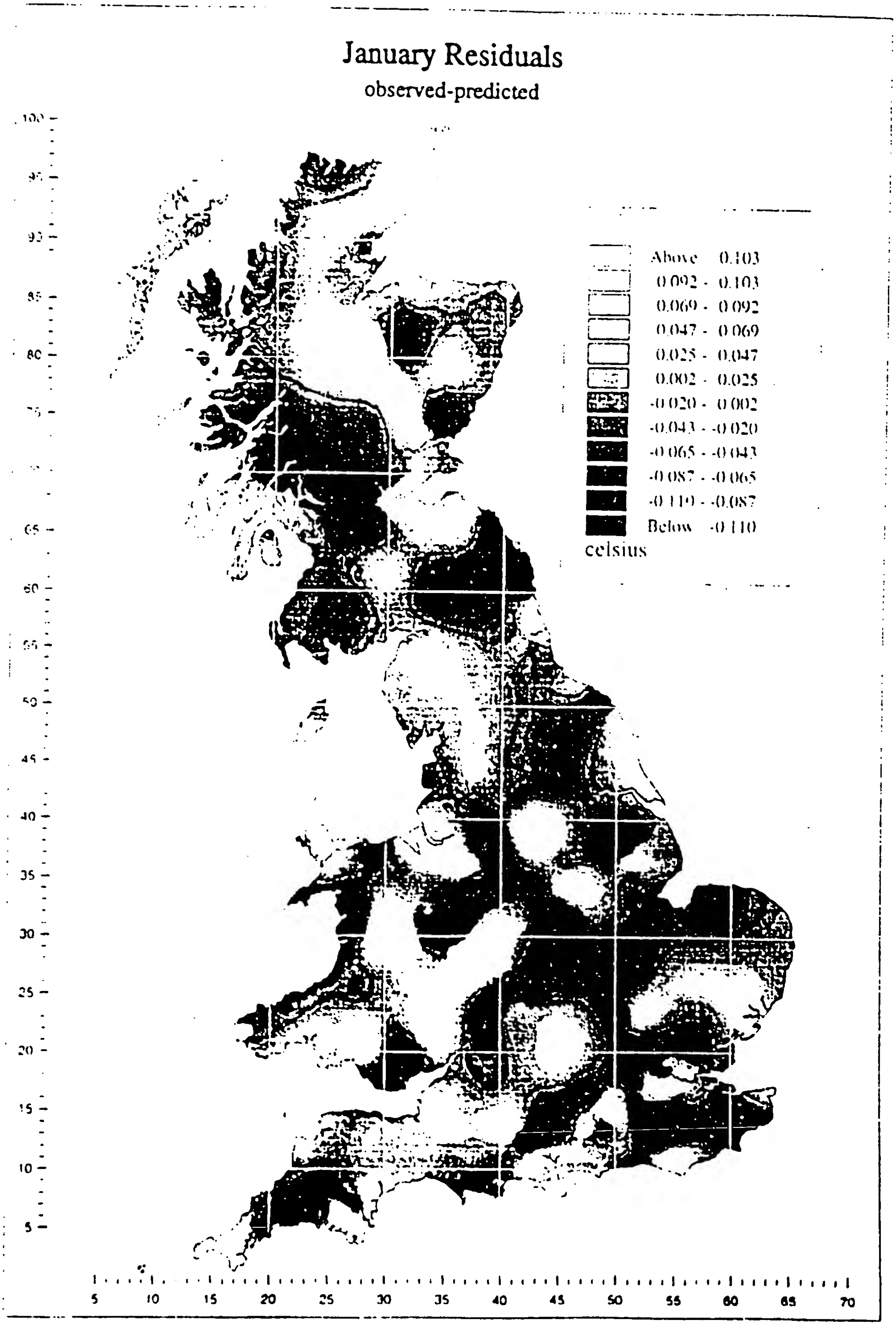

January Residuals
observed-predicted
Above 0.103
0.092 - 0.103
0.069 - 0.092
0.047 - 0.069
0.025 - 0.047
0.002 - 0.025
-0.020 - 0.002
-0.043 - -0.020
-0.065 - -0.043
-0.087 - -0.065
-0.110 - -0.087
Below -0.110
celsius
5
10
15
20
25
30
35
40
45
50
55
60
65
70

Figure 5

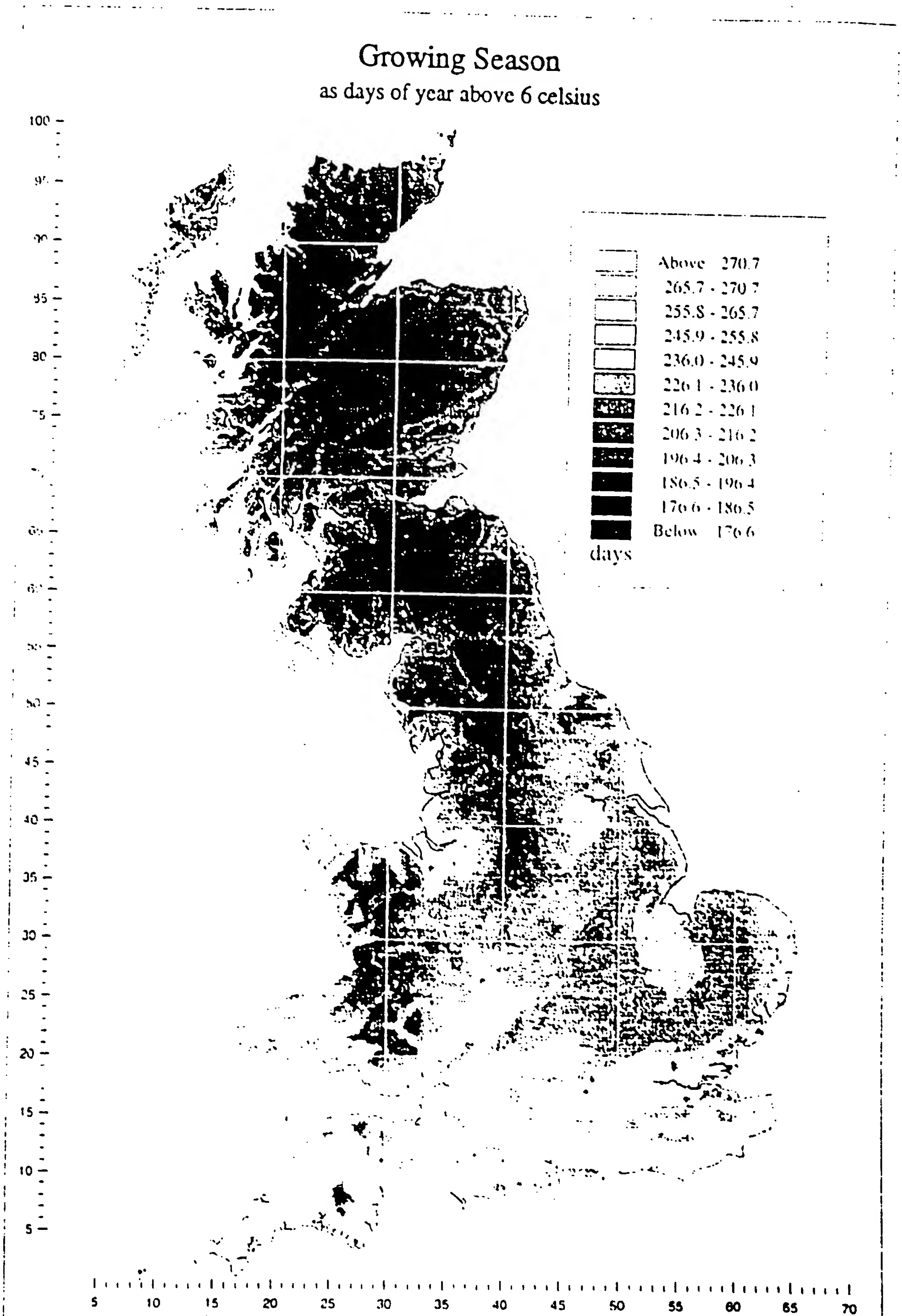
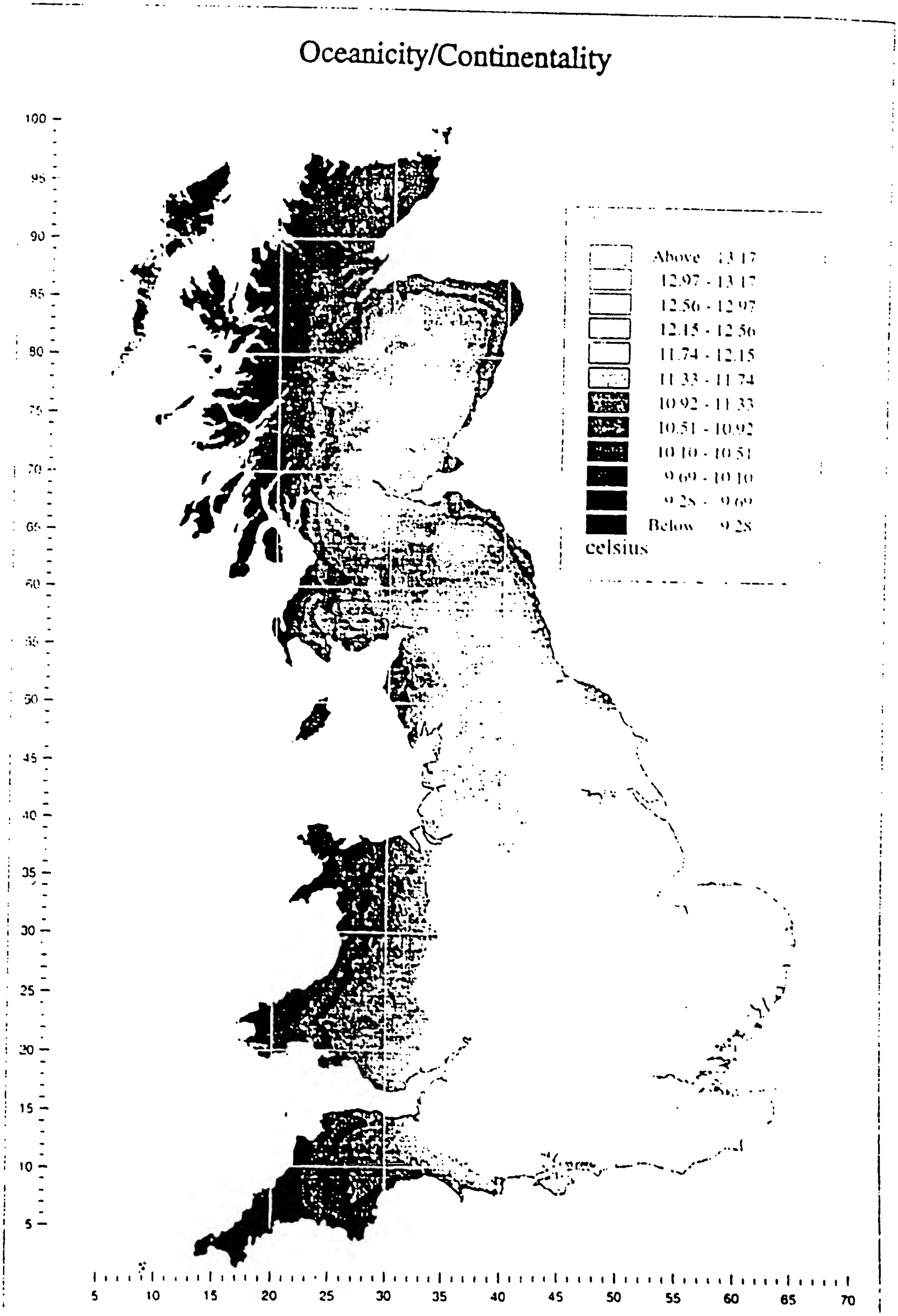

Growing Season
as days of year above 6 celsius
Above 270.7
265.7 - 270.7
255.8 - 265.7
245.9 - 255.8
236.0 - 245.9
226.1 - 236.0
216.2 - 226.1
206.3 - 216.2
196.4 - 206.3
186.5 - 196.4
176.6 - 186.5
Below 176.6
days
100
85
80
75
45
40
35
30
25
20
15
10
5
5
10
15
20
25
30
35
40
45
50
55
60
65
70

Figure 6


Oceanicity/Continentality
Above 13.17
12.97 - 13.17
12.56 - 12.97
12.15 - 12.56
11.74 - 12.15
11.33 - 11.74
10.92 - 11.33
10.51 - 10.92
10.10 - 10.51
9.69 - 10.10
9.28 - 9.69
Below 9.28
celsius
100
95
90
85
80
75
70
65
60
55
50
45
40
35
30
25
20
15
10
5
5
10
15
20
25
30
35
40
45
50
55
60
65
70

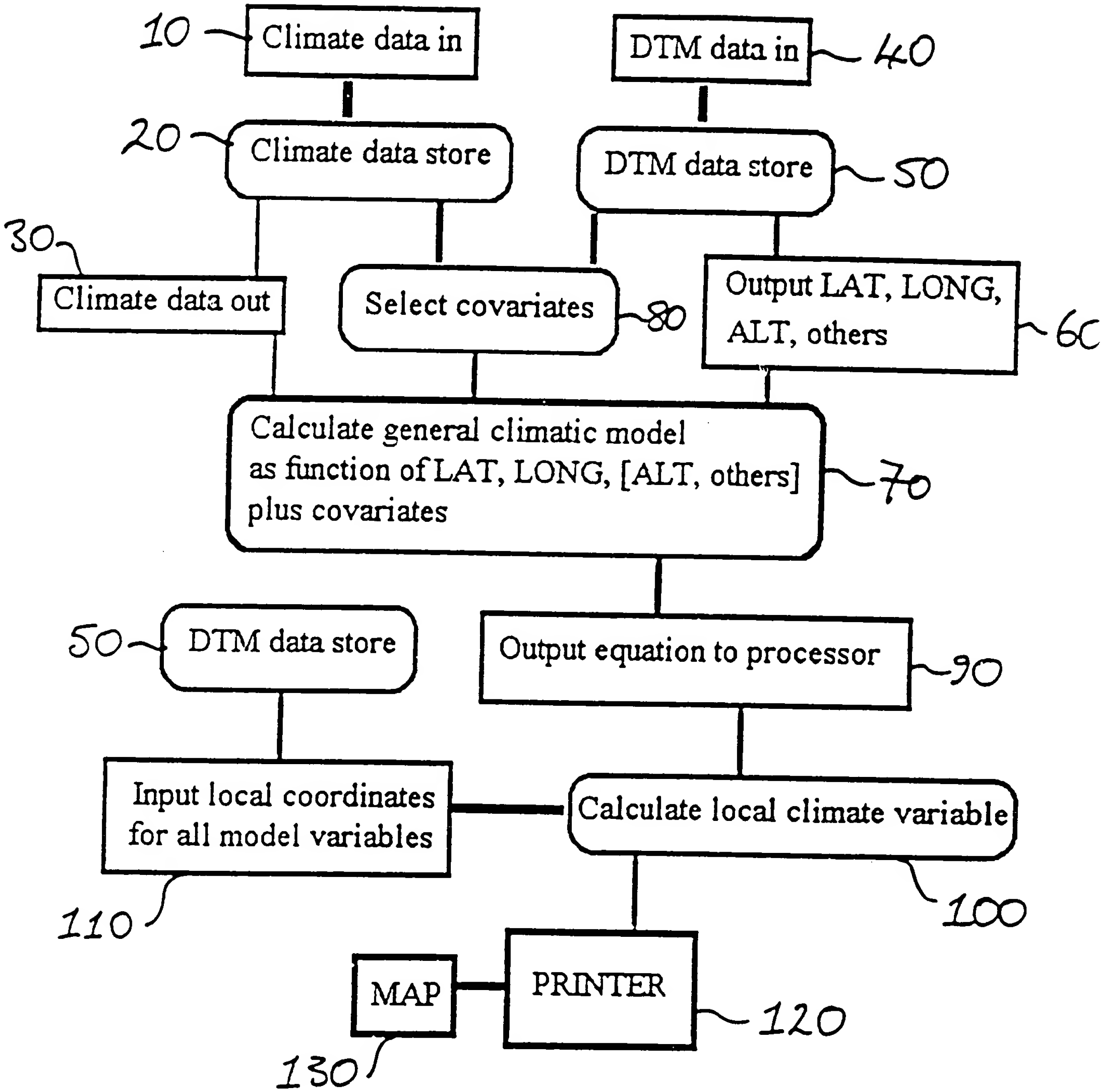

10
Climate data in
DTM data in
40
20
Climate data store
DTM data store
50
30
Climate data out
Select covariates
80
Output LAT, LONG, ALT, others
60
Calculate general climatic model as function of LAT, LONG, [ALT, others] plus covariates
70
50
DTM data store
Output equation to processor
90
Input local coordinates for all model variables
Calculate local climate variable
110
100
MAP
PRINTER
120
130

FIG.7

INTERNATIONAL SEARCH REPORT

International Application No
PCT/GB 96/00700

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 G01W1/10 G09B29/00

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 G01W G06F G09B

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	11: "PROCEEDINGS OF IGARSS '86 SYMPOSIUM,8-11 SEPT.1986" August 1986 , ESA , ZÜRICH XP002005675 vol.3,G.MENZ:"Deduction of a Synthetic Bioclimatological Map by Means of...",pages 1525-1529 see the whole document ---	1-3
A	RADIO FERNSEHEN ELEKTRONIK, vol. 34, no. 2, 1986, BERLIN, pages 117-121, XP002005674 HANS STEINHAGEN: "Automatisches Erfassungs-und Verarbeitungssystem AES-2 für Radiosondenmessungen" see the whole document ---	1
P,A	WO,A,95 13547 (UNISYS CORP) 18 May 1995 see the whole document ---	1,2

-/--

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
14 June 1996	02.07.96

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax (+31-70) 340-3016	Gorun, M

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US,A,5 221 924 (WILSON JR F WESLEY) 22 June 1993 see the whole document -----	1,2

PCT/GB 96/00700

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9513547	18-05-95	US-A- 5451961	19-09-95
		AU-B- 7877594	29-05-95
US-A-5221924	22-06-93	AU-B- 645036	06-01-94
		AU-B- 6454990	18-04-91

This Page is Inserted by IFW Indexing and Scanning Operations and is not part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☑ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: ______________________

IMAGES ARE BEST AVAILABLE COPY.
As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.